A Deep Reinforcement Learning PI Tuning Strategy for Closed-loop Operation of a Recirculating Aquaculture System

\*Corresponding author. E-mail address: [laricard@uwaterloo.ca](mailto:laricard@uwaterloo.ca)

Abstract

Adequate water quality is key for fish production in recirculating aquaculture systems (RAS). In this study, we formulate the control of RAS water quality parameters using a deep Reinforcement Learning (RL) based multi-loop proportional-integral (PI) tuning strategy. The key novelty is the design of an adaptive RL agent to respond to the nonlinear and highly interactive behaviour of the multiple-input, multiple-output RAS in real-time while handling constraints for water quality parameters and sensor malfunctions. We tested our tuning strategy in a scenario where multiple sensors can malfunction simultaneously. The proposed RL agent significantly outperformed conventional PI controllers in oxygen concentration setpoint tracking while reducing critical and supersaturation dissolved oxygen level violations. The setpoint tracking performances of total ammonia and nitrate concentrations were slightly sacrificed, but this did not compromise fish health as these concentrations remained well below toxic levels. The RL-based PI tuning strategy exhibits acceptable control performance for water quality parameters and sheds light on future applications of RL-based closed-loop RAS operation.

**Keywords**: Recirculating aquaculture systems, Reinforcement Learning, PI controllers

* 1. Introduction

Recirculating aquaculture systems (RAS) are attractive options for fish production due to their substantial benefits, including decreased water usage, environmentally friendly, and improved productivity (Kamali et al., 2022). Adequate control of RAS in terms of water quality is crucial for fish growth (Kamali et al., 2023). While conventional proportional-integral-derivative (PID) controllers with fixed tuning parameters remain the most widely adopted and popular controllers in the industry, they may not perform well for multiple-input, multiple-output (MIMO) RAS given the nonlinear and highly interactive behaviour of these systems. For instance, conventional PID controllers cannot account for process constraints and real-time response to major changes during operation, such as sensor malfunction. Hence, studies involving adaptive PID controllers that maintain RAS operation on target have emerged. Zhou et al. (2021 & 2022) proposed a differential evolution algorithm-optimized radial basis function neural network PID controller and a fuzzy rule-optimized single-neuron adaptive PID controller for dissolved oxygen (DO) control in RAS, respectively. Those studies only considered a single-input, single-output RAS and did not include water quality constraints. In recent years, studies of Reinforcement Learning (RL) in adaptive PID control have become popular. For instance, Carlucho et al. (2020) proposed a deep RL-based adaptive MIMO PID tuning approach using an inverted deep deterministic policy gradient (IDDPG) algorithm to control mobile robots. Yu et al. (2022) constructed a model-free self-adaptive SAC-PID control approach for mobile robots using the soft actor-critic algorithm.

This study formulates a deep RL-based proportional-integral (PI) tuning approach for controlling water quality parameters under constraints, i.e., oxygen and total ammonia (TAN), in MIMO RAS subjected to water quality sensor malfunctions, which are aspects that have not been addressed in the literature; hence, the novelty of the proposed framework. The proposed control scheme is an adaption of the IDDPG algorithm proposed by Carlucho et al. (2020) with a tailored RL agent design. To the authors’ knowledge, the present study is the first that considers RL for closed-loop RAS operation.

Problem Statement

In this study, the dynamic mechanistic model proposed by Kamali et al. (2022) is adopted to predict the transient behaviour of RAS. The model consists of a fish-rearing tank, a mechanical filter, and two moving bed biofilm reactors (BRs) to treat the water leaving RAS. The model mainly involves mass balances of waste components and oxygen. There are 14 waste components () in the RAS model, including nitrate nitrogen and TAN. The fish-rearing tank is modelled as a well-mixed reactor. The mass balance of waste components’ concentrations () in the fish tank is as follows:

|  |  |
| --- | --- |
|  | (1) |

where refers to either the concentration of the soluble () or particulate () waste component and is the corresponding inlet concentration; and are the volume and flow rate to the fish tank, respectively. and are the waste production rate and feed loss of waste component , respectively, i.e. TAN and nitrate,

|  |  |
| --- | --- |
|  | (2) |

where and are the fractions of component in waste and feed, respectively. , , and are the feed residence time, feed loss fraction, and feeding rate, respectively. It is assumed that fish growth is slow and negligible compared to changes in the waste components and oxygen concentrations. Also, the removal rate of the mechanical filter downstream of the fish-rearing tank is assumed constant and operated at steady-state.

In this study, BRs are modelled using a zero-dimensional biofilm model. Thus, the biochemical conversions in the bulk and biofilm for suspended particulate (), dissolved (), and biofilm particulate () waste components are described as follows:

|  |  |
| --- | --- |
|  | (3) |

where and are the volume and flow rate to BRs, respectively; represents , , or ; is the inlet concentration; is the rate of change of component .

The oxygen concentration mass balance is the same for the fish-rearing tank (FT) and BRs. Thus, for , the expression is as follows:

|  |  |
| --- | --- |
|  | (4) |

where , , , and are the oxygen’s bulk concentration, inlet concentration, consumption rate, and addition rate, respectively; is the volume and is the flow rate. The complete list of waste components () and additional modelling details can be found in Kamali et al. (2022).

The controlled variables and manipulated variables available to maintain a suitable environment for fish rearing are , , and , , , respectively. , , and are the oxygen, TAN, and nitrate concentrations in the fish tank, respectively; denotes the make-up water flow rate to the fish tank. Through a Relative Gain Array analysis (not shown for brevity), three control loop pairings were identified for this process, i.e., , , and . The control objective is to use the three control loops to change values in to maintain close to a setpoint . Since consists of water quality parameters, it is desired to keep within known fish species-dependent ranges for optimal fish growth, i.e., , 0, 0] and . , , , and are the critical DO, supersaturation DO, TAN toxic, and nitrate toxic levels, respectively. The multiloop PI control scheme is constructed as follows. For each control loop , a sensor and a PI controller with proportional gain () and integral time constant () are considered. The tuning parameters are denoted as ; for .

According to a report from the Government of Newfoundland & Labrador (2014), water quality sensors are prone to calibration and biofouling drifts. This condition has not been widely explored for MIMO RAS. In this work, we consider a potential sensor malfunction in the three water quality sensors measuring with measurements. It is assumed that the deployment period for all sensors is days. During this period, each sensor may start to malfunction on Days , , or with respective probabilities , , and . That is, a sensor would start to malfunction on Day and persist until the end of the -day period if , where is a random number sampled from a uniform distribution, i.e., . This procedure is repeated if the sensor is operating normally on Days or . At the end of the -day period, all sensors are re-calibrated and ready for the next -day deployment period. The realization of the sensor malfunction is in the form of a zero-mean Gaussian noise () with a standard deviation of sensor measurement added to the actual sensor measurement ().

The objective of the problem constructed in this study is to design multi-loop PI controllers for the RAS plant model presented above under sensor malfunctions and water quality parameter constraints by tuning the PI controller parameters . Typically, is tuned offline using conventional PI tuning methods such as Internal Model Control (IMC). However, such tuning approaches usually lead to poor control performance due to a lack of real-time response to the nonlinear and highly interactive behaviour of MIMO RAS in the presence of sensor malfunctions and water quality parameter constraints. To tackle this problem, this work presents an adaptive RL-based PI tuning approach which is described in the next section.

Mathematical Framework

Our mathematical framework was adapted from the IDDPG algorithm proposed by Carlucho et al. (2020), as it allows a wide continuous action space range () and constraints output actions within bounds and prevents saturation by inverting the critic’s gradients. IDDPG contains the basic elements for an RL agent, i.e., state , action , and reward . Also, since IDDPG is a model-free, off-policy, actor-critic algorithm, it also consists of an actor-network and a critic-network with weights and , target networks and with weights and to soft update the learned networks, and a replay buffer to store transitions to be used for RL agent training. Details of IDDPG can be found in Carlucho et al. (2020). A key novelty in our framework involves designing our own action vector and reward function for the RL agent. Our framework also considers RAS sensor malfunctions in the RL agent training. For an episode with maximum simulation time and sampling time , the selection of at a sampling time step is expressed as follows:

|  |  |
| --- | --- |
|  | (5) |

where and are the upper and lower bounds of . Unlike Carlucho et al. (2020), we utilize noise sampled from a Gaussian noise process with mean , standard deviation , and an identity matrix to encourageexploration. Also, we apply a logistic action noise scaling factor with constants and for the scheduled reduction of action noise with increases in the number of episodes to improve the robustness of the learning process in our framework. Through executing , sensor measurements are recorded. For any sensor that malfunctions, a sensor noise is added to such that . The sensor malfunction is then introduced into the RL agent training by constructing the next state with , , and the error vector . The reward function makes use of the sum of squares error (SSE) for setpoint tracking while penalizing constraint violations in and , i.e.,

|  |  |
| --- | --- |
|  | (6) |

The episodic return is thus calculated as ER=. Note that subscript and superscript above refer to the sampling time step and controller/sensor, respectively.

* 1. Results and Discussions

Case study

The RL strategy presented in the previous section was tested on a case study of the MIMO RAS plant model presented in Kamali et al. (2023). The details of the RAS described in Section 2 are as follows. The volumes of the fish-rearing tank () and two BRs (, ) are 5.5 , 0.4 , and 0.4 , respectively; a 40 drum filter is used as the mechanical filter. The RAS is used to raise rainbow trout at with a constant body weight of 0.045 kg. The feed schedule is 6 hours per day with a 3 kg/day feed rate. 0.008, 0.00062, 0.0209] kg/, 0.007, 0, 0] kg/, and 0.013, 0.013, 0.075] kg/. The rest of the model parameters and nominal conditions can be found in Kamali et al. (2023). Moreover, for sensor malfunctions, a common deployment period of days is considered; , , and are set to Days 1, 11, and 21. For all , = 10 %, = 30 %, and = 50 %, respectively; also, % = 10 %. For comparison, the three PI controllers were also tuned using IMC followed by a manual adjustment for better control performance. The resulting PI tuning parameters are 14.5, 0.3], [-8945.437, 0.15], [-500, 0.8]]. and in Eq. (5) are defined as [25, 5], [-1000, 5], [-80, 5]] and [0.1, 0.02], [-9000, 0.02], [-800, 0.02]].

RL agent settings, training, and validation

In terms of RL agent settings, the parameter values used in this study for the Gaussian noise and the logistic schedular in action exploration: , , , and were set to , 0.01, 0.02, and 250, respectively. For the RL agent training, was set to days to accommodate the timescale differences and daily fluctuations in RAS. To train the agent with diverse sensor malfunction instances, was set to 90 days to allow three 30-day deployment periods () within an episode. The agent underwent 1,500 training episodes on a PC with Intel® Core™ i7-9700K CPU @ 3.60 GHz and 64 GB of RAM using TensorFlow 1 in Python. The RAS model was simulated using an interior-point solver.

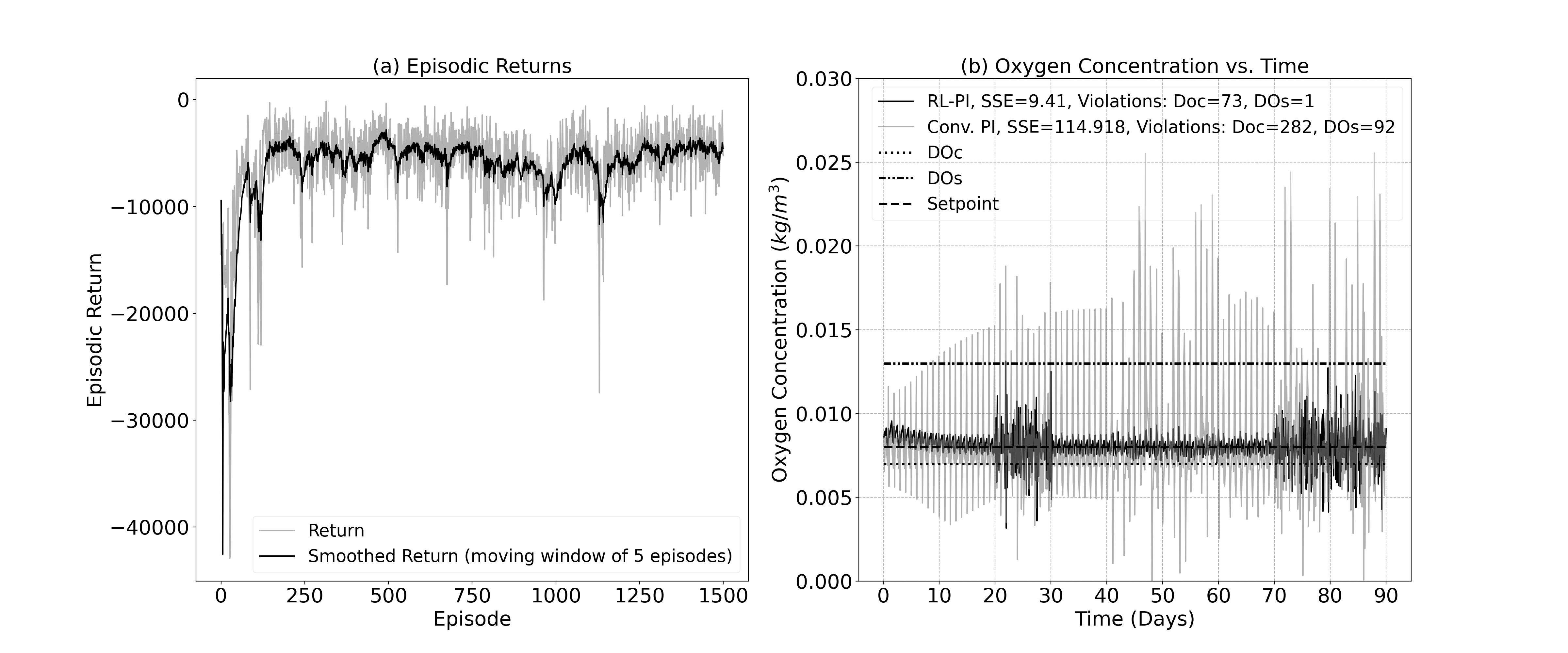


Figure 1. (a) Episodic returns and smoothed returns of the RL agent; (b) Oxygen concentration setpoint tracking performances of both control methods under one sensor malfunction instance.

As shown in Figure 1a, although some fluctuations were observed at the beginning of the learning curve, the agent’s training converged after around 250 episodes. The setpoint tracking performances of the agent were validated using 200 testing runs with different sensor malfunction instances. A comparison was made with in terms of the means and standard deviations of SSEs and constraint violation counts for , , , and . For better visualization, SSEs were magnified by a factor of . As depicted in Table 1, the agent significantly outperforms in setpoint tracking and reducing violations in and . Compared to , the agent reduced the means of SSE of and violation count by 2 orders of magnitude and the mean of violation count by 1 order of magnitude. Nevertheless, these good performances came at the expense of slightly sacrificing and setpoint tracking performances. That is, the agent produced actions that resulted in the means of SSEs of and that are 0.80 % and 14.36 % higher than those obtained for , respectively. However, fish health was not compromised as and remained below toxic levels, i.e., no violations in and were observed. Regarding standard deviations, compared to , the agent reduced the variability in most SSEs and violation counts, except for the SSE of and violation count. This indicates that for the agent, the different combinations of sensors that may malfunction during operation can create variability in setpoint tracking performance and violation count, and may lead to cases of large SSE of and high violation count.

The oxygen concentration setpoint tracking performances for both the RL agent and conventional PI controllers under one sensor malfunction instance from the 200 testing runs are illustrated in Figure 1b. For this instance, the oxygen, TAN, and nitrate concentration sensors malfunction on Days 21~30 and 71~90, Days 41~60 and 71~90, and Days 21~30 and 61~90, respectively. As shown in Figure 1b, for the RL agent, and constraint violations only occur when there are oxygen concentration sensor malfunctions (Days 21~30 and 71~90), which is reflected by the large variability observed in this parameter. For the rest of the simulation time, the RL agent is able to track the setpoint and avoid constraint violations, even in the presence of other sensor malfunctions. Conversely, conventional PI controllers continuously violate and limits with large variability due to sensor malfunctions.

Table 1. Comparison of setpoint tracking performances in the form of mean and standard deviation (mean deviation) of SSEs and constraint violation counts.

|  |  |  |
| --- | --- | --- |
| Controlled Variables () | SSE: RL-PI | SSE: PI |
| Oxygen Concentration () |  |  |
| TAN Concentration () |  |  |
| Nitrate Concentration () |  |  |
| Constraints | Violation Count: RL-PI | Violation Count: PI |
| DO Critical () |  |  |
| DO Supersaturation () |  |  |
| Toxic Levels & |  |  |

Conclusions

This study presents a mathematical framework of a deep RL-base tuning strategy to solve the multi-loop PI tuning problem for the control of RAS water quality parameters (i.e., oxygen, TAN, and nitrate concentrations) with fish species-dependent constraints and subject to multiple water quality sensor malfunctions. The key novelties of this work are the application of deep RL to closed-loop RAS operation, improvements to the IDDPG algorithm by RL agent design, and incorporation of water quality sensor malfunction into RL agent training. We applied the proposed strategy to the closed-loop operation of RAS, where multiple sensors could malfunction simultaneously. The results showed that the proposed RL strategy exhibits acceptable control performances for water quality parameters and significantly reduced critical and supersaturation DO level violations. Thus, this RL strategy has the potential to be implemented online for closed-loop RAS operations to maintain water quality to their setpoints while reducing potential risks to fish health. Future work includes improving the proposed strategy by incorporating fish growth dynamics into RAS and optimizing for fish growth and water quality control.

References

I. Carlucho, M. De Paula, & G. G. Acosta, 2020, An adaptive deep reinforcement learning approach for MIMO PID control of mobile robots, ISA transactions, 102, 280-294.

S. Kamali, V. C. A. Ward, & L. Ricardez-Sandoval, 2022, Dynamic modeling of recirculating aquaculture systems: Effect of management strategies and water quality parameters on fish performance, Aquacultural Engineering, 99, 102294.

S. Kamali, V. C. A. Ward, & L. Ricardez-Sandoval, 2023, Closed-loop operation of a simulated recirculating aquaculture system with an integrated application of nonlinear model predictive control and moving horizon estimation, Computers and Electronics in Agriculture, 209, 107820.

Water Resources Management Division, Department of Environment and Conservation, Government of Newfoundland & Labrador, 2014, Protocols manual for RTWQ monitoring in NL – calibration and maintenance guide for industry partners, https://www.gov.nl.ca/ecc/files/waterres-rti-rtwq-nl-rtwq-manual-calibration.pdf

X. Yu, Y. Fan, S. Xu, & L. Ou, 2022, A self‐adaptive SAC‐PID control approach based on reinforcement learning for mobile robots, International journal of robust and nonlinear control, 32(18), 9625-9643.

X. Zhou, D. Li, L. Zhang, & Q. Duan, 2021, Application of an adaptive PID controller enhanced by a differential evolution algorithm for precise control of dissolved oxygen in recirculating aquaculture systems, Biosystems Engineering, 208, 186-198.

X. Zhou, J. Wang, L. Huang, D. Li, & Q. Duan, 2022, Modelling and controlling dissolved oxygen in recirculating aquaculture systems based on mechanism analysis and an adaptive PID controller, Computers and Electronics in Agriculture, 192, 106583.